B8 cont'd

PCT/US01/00663; PCT/US01/00662; PCT/US01/00661; and PCT/US01/00670, the disclosures of which are incorporated herein by reference in their entireties.

IN THE CLAIMS:

Please add the following new claims 21 - 92:

B9

21 (new). A high throughput, microarray-based method to confirm predicted exons, comprising:

detecting hybridization by transcript-derived nucleic acids to microarray probes that include genomic sequence predicted to contribute to no more than one exon,

detectable hybridization confirming the prediction of the exon included in each of said detectably hybridized probes.

22 (new). The method of claim 21, wherein at least 75% of the probes of said microarray include genomic sequence predicted to contribute to no more than one exon.

23 (new). The method of claim 21, wherein at least 90% of the probes of said microarray include genomic sequence predicted to contribute to no more than one exon

24 (new). The method of claim 21, wherein at least 95% of the probes of said microarray include genomic sequence predicted to contribute to no more than one exon.

25 (new). The method of claim 21, wherein said genomic sequence is human genomic sequence.

B9 cont'd

26 (new). The method of claim 21, wherein said prediction is output from a computer program selected from the group consisting of GenScan, Diction, Genefinder, and Grail.

27 (new). The method of claim 26, wherein said prediction is output from GenScan.

28 (new). The method of claim 21, wherein said microarray has probes that collectively include exons predicted from all chromosomes of a eukaryotic organism.

29 (new). The method of claim 28, wherein said eukaryotic organism is a human being.

30 (new). The method of claim 21, wherein said microarray has probes that include exons predicted from human chromosome 22.

31 (new). The method of claim 21, wherein each of said predicted exons is represented by a plurality of probes on said array.

32 (new). The method of claim 21, wherein said microarray includes between 5,000 and 19,000 probes.

33 (new). The method of claim 21, wherein the genomic sequence included within said probes is selected at least in part based upon considerations of base composition and/or hybridization binding stringency.

34 (new). The method of claim 21, wherein said probes include at least 50 nt of predicted exon.

35 (new). The method of claim 21, wherein said probes include at least 75 nt of predicted exon.

36 (new). The method of claim 21, wherein said probes are amplified from genomic DNA.

37 (new). The method of claim 21, wherein said probes are chemically synthesized.

B9 Cont'd

38 (new). The method of claim 21, wherein said probes are noncovalently attached to the substrate of said microarray.

39 (new). The method of claim 21, wherein said probes are covalently attached to the substrate of said microarray.

40 (new). The method of claim 21, wherein said probes are disposed on said microarray substrate by ink jet.

41 (new). The method of claim 21, wherein the substrate of said microarray is a glass slide.

42 (new). The method of claim 21, further comprising the antecedent step of:

contacting said microarray with at least a first sample of transcript-derived nucleic acids, said nucleic acids being detectably labeled.

43 (new). The method of claim 42, wherein said transcript-derived nucleic acids are first strand cDNA.

44 (new). The method of claim 43, wherein said cDNAs are fluorescently labeled.

45 (new). The method of claim 44, wherein said fluorescent label is selected from the group consisting of Cy3 and Cy5.

46 (new). The method of claim 42, wherein said contacting step comprises contacting said microarray concurrently with a first sample of transcript-derived nucleic acids and with a second sample of transcript-derived nucleic acids, wherein said first and second samples are labeled respectively with a first and a second label, said first and second labels being separately detectable.

B9 Cont'd

47 (new). The method of claim 46, wherein said detecting includes normalizing and background correcting signals from each of said labels.

48 (new). The method of claim 46, wherein said labels are Cy3 and Cy5.

49 (new). The method of claim 46, wherein said first sample includes transcript-derived nucleic acids pooled from a plurality of tissues and/or cell types.

50 (new). The method of claim 49, wherein said pool includes transcript-derived nucleic acids from a plurality of human cell lines.

51 (new). The method of claim 49, wherein the transcript-derived nucleic acids of said second sample are derived from a cell line or normal tissue.

52 (new). The method of claim 51, wherein the transcript-derived nucleic acids of said second sample are derived from a source within the group of human tissues and

cell lines consisting of: brain, heart, liver, fetal liver, placenta, lung, bone marrow, HeLa cells, BT474 cells and HBL 100 cells.

53 (new). A method of identifying potential false positive exon predictions, comprising:

detecting hybridization by transcript-derived nucleic acids to a microarray that has probes that include genomic sequence predicted to contribute to no more than one exon,

absence of detectable hybridization identifying as a potential false positive the exon predicted in each undetectably hybridized probe.

54 (new). A method of identifying one or more genes expressed by one or more eukaryotic cells having a genome that averages at least one intron per gene, comprising:

(a) contacting a cDNA sample prepared by enzymatically copying messenger RNA obtained from said eukaryotic cell(s) into cDNA, wherein said cDNA comprises a detectable label, with a plurality of single exon probes, each said single exon probe comprising a discrete nucleic acid sequence encoding all or a portion of a single exon of said eukaryotic genome that specifically hybridizes at high stringency to a target nucleic acid when said target nucleic acid is present in said cDNA sample;

(b) detecting a signal from each said single exon probe that is specifically hybridized to said target nucleic acid, wherein the presence of said signal indicates the expression of a gene comprising said single exon by said eukaryotic cell(s).

55 (new). A method of identifying one or more genes expressed by one or more human cells, comprising:

(a) contacting a cDNA sample prepared by copying messenger RNA obtained from said human cell(s) into cDNA using reverse transcriptase, wherein said cDNA comprises a detectable label, with a nucleic acid microarray, said microarray comprising a substantially planar glass substrate comprising (i) at least 5000 addressable locations to which single exon probes are bound, each said single exon probe comprising a discrete nucleic acid sequence encoding all or a portion of a single exon of a human genome that is specifically hybridizable at high stringency to a target nucleic acid, wherein said target nucleic acid is a sequence encoding all or a portion of an expressed gene, or a complementary sequence thereof, and (ii) one or more additional locations to which control nucleic acid sequences are bound; and

B9 cont'd

(b) generating a signal from each said addressable location, wherein the presence of a signal at a specific addressable location indicates the expression by said human cell(s) of a gene comprising the single exon probe bound to that addressable location.

56 (new). A high throughput, microarray-based method of grouping exons into a common gene, comprising:

comparing the patterns of tissue and/or cell-type expression of exons predicted from a contiguous region of genomic DNA,

wherein said patterns of expression have been determined by detecting hybridization of transcript-derived nucleic acids from a plurality of tissues and/or cell types to microarray probes, each of said probes including genomic sequence predicted to contribute to no more than one of said exons, said microarray including probes that collectively comprise all of said exons,

consensus in said expression patterns identifying exons that are groupable into a common gene.

B9 Cont-d

57 (new). The method of claim 56, wherein said gene is a human gene.

58 (new). The method of claim 56, wherein said patterns are detected by detecting (i) fluorescence intensity, (ii) the ratio of intensity as between concurrently hybridized first and second samples, or (iii) a combination of (i) and (ii).

59 (new). A nucleic acid microarray comprising:

a substrate comprising a plurality of addressable locations to which nucleic acid sequences are bound; and

a plurality of single exon probes bound at said addressable locations, each said single exon probe comprising a discrete nucleic acid sequence encoding all or a portion of a single exon of a eukaryotic genome averaging at least one intron per gene that is specifically hybridizable at high stringency to a target nucleic acid, wherein said target nucleic acid is a sequence encoding all or a portion of an expressed gene, or a complementary sequence thereof.

60 (new). A nucleic acid microarray comprising:

a substantially planar glass substrate comprising (i) at least 5000 addressable locations to which single exon probes are bound, each said single exon probe comprising a discrete nucleic acid sequence encoding all or a portion of a single exon of a human genome that is specifically hybridizable at high stringency to a target nucleic acid, wherein said target nucleic acid is a sequence encoding all or a portion of an expressed gene, or a complementary

sequence thereof, and (ii) one or more additional locations to which control nucleic acid sequences are bound.

61 (new). A single exon nucleic acid microarray, comprising:

B9 cont'd

a plurality of nucleic acid probes addressably disposed upon a substrate,

wherein at least 50% of said probes include genomic sequence predicted to contribute to no more than one exon of a eukaryotic genome, said eukaryotic genome averaging at least one intron per gene, and wherein said plurality of nucleic acid probes averages at least 50 nt in length.

62 (new). The microarray of claim 61, wherein at least 75% of said nucleic acid probes include genomic sequence predicted to contribute to no more than one exon of a eukaryotic genome.

63 (new). The microarray of claim 61, wherein at least 90% of the probes of said microarray include genomic sequence predicted to contribute to no more than one exon of a eukaryotic genome.

64 (new). The microarray of claim 61, wherein at least 95% of the probes of said microarray include genomic sequence predicted to contribute to no more than one exon of a eukaryotic genome.

65 (new). The microarray of claim 61, wherein said microarray has probes that collectively include exons predicted from all chromosomes of a eukaryotic genome.

66 (new). The microarray of claim 61, wherein said eukaryotic genome is a human genome.

67 (new). The microarray of claim 65, wherein said eukaryotic genome is a human genome.

68 (new). The microarray of claim 61, wherein said prediction is output from a computer program selected from the group consisting of GenScan, Diction, Genefinder, and Grail.

B9 cont'd

69 (new). The microarray of claim 68, wherein said prediction is output from GenScan.

70 (new). The microarray of claim 61, wherein each of said predicted exons is represented by a plurality of probes on said array.

71 (new). The microarray of claim 61, wherein said microarray includes between 5,000 and 19,000 probes.

72 (new). The microarray of claim 61, wherein the genomic sequence included within said probes is selected at least in part based upon considerations of base composition and/or hybridization binding stringency.

73 (new). The microarray of claim 61, wherein said probes have been amplified from genomic DNA.

74 (new). The microarray of claim 61, wherein said probes have been chemically synthesized.

75 (new). The microarray of claim 61, wherein said probes are noncovalently attached to the substrate of said microarray.

76 (new). The microarray of claim 61, wherein said probes are covalently attached to the substrate of said microarray.

77 (new). The microarray of claim 61, wherein said probes are disposed on said microarray substrate by ink jet.

B7 cont'd

78 (new). The microarray of claim 61, wherein said substrate is a glass slide.

79 (new). The microarray of claim 61, wherein each of said probes is disposed on said array with its reverse complement.

80 (new). The microarray of claim 61, further comprising control probes.

81 (new). The microarray of claim 61, wherein at least 50% of said exon-including nucleic acid probes comprise, contiguous to a first end of said predicted exon, a first intronic and/or intergenic sequence that is identically contiguous to said exon in the human genome, and further comprise, contiguous to a second end of said predicted exon, a second intronic and/or intergenic sequence that is identically contiguous to said exon in the human genome.

82 (new). A software data structure for annotating nucleic acid sequence with confirmed bioinformatic predictions, the data structure stored in a machine readable medium and comprising:

a plurality of sequence entries, each sequence entry including (i) a sequence identifier and (ii) software means for relating said sequence identifier to data that

encode a confirmed prediction of a biological function of the nucleic acid sequence identified by said sequence identifier.

83 (new). The software data structure of claim 82, wherein said confirmed biological function is contribution to a mature mRNA transcript.

B9 cont'd

84 (new). The software data structure of claim 83, wherein said prediction is output from GenScan.

85 (new). The software data structure of claim 83, wherein said prediction has been confirmed by the method of claim 21.

86 (new). The software data structure of claim 82, wherein said software relating means is the common inclusion of said confirmed prediction data in a single record with said sequence identifier.

87 (new). The software data structure of claim 82, wherein said software relating means links said sequence identifier to confirmed prediction data present in a distinct record.

88 (new). The software data structure of claim 82, wherein said sequence entries further comprise:

software means for relating said sequence identifier to data that encode at least one nucleic acid sequence identified by said identifier.

89 (new). The software data structure of claim 88, wherein said sequence entries further comprise:

software means for relating said sequence identifier and/or said at least one nucleic acid sequence to

data that encode a measure of similarity of the at least one nucleic acid sequence to at least one nucleic acid sequence prior-accessioned into a database.

90 (new). The software data structure of claim 89, wherein said sequence entries further comprise:

software means for relating said sequence identifier and/or said at least one nucleic acid sequence to data that encode a textual description of said at least one similar prior-accessioned nucleic acid sequence.

91 (new). The software data structure of claim 82, wherein said sequence entries further comprise:

software means for relating said sequence identifier to data that encode a chromosomal map location of the sequence identified by said sequence identifier.

92 (new). An isolated nucleic acid having exons that have been commonly grouped by the method of claim 56.

REMARKS

Amendments to the Specification

(1) Applicants have amended the specification by replacing the paragraph that begins at line 19 on page 5 and ends at line 5 of page 6 in order to correct a clerical error in the spelling of the gene finding program GENSCAN (from "GENESCAN"). The correct spelling appears in the specification as filed at page 25, line 28; page 27, line 23; p. 65, lines 30 - 31; and p. 94, line 5; accordingly, no new matter has been added.